Bayesian linear regression*

February 19, 2021

1 Learning about parameters

We focus here on learning unknown parameters of a linear regression model. The unknown parameters in this case can be interpreted as *hidden constant* states. The unknown parameters are therefore a source of invariant events and conditioning on the parameters is equivalent to conditioning on the invariant events. We will use (Bayesian) conjugate prior analysis of Luce and Raiffa (1957) to learn about parameters sequentially from increments to a data history. Specifically, we assign a prior to the parameters, and recursively update the priors as as we observe more data.

Consider the following VAR:

$$\begin{aligned}
\mathbf{X}_{t+1} &= \mathbf{A}\mathbf{X}_t + \mathbf{B}\mathbf{W}_{t+1} \\
\mathbf{Y}_{t+1} - \mathbf{Y}_t &= \mathbf{H} + \mathbf{D}\mathbf{X}_t + \mathbf{F}\mathbf{W}_{t+1}, \\
\mathbf{W}_{t+1} &\overset{\text{iid}}{\sim} N(0, \mathbb{I}).
\end{aligned}$$

with an observable state vector $\mathbf{X}_t$ but unknown parameter $\mathbf{H}$. We suppose $\mathbf{F}$ is nonsingular (which implies that $\mathbf{Y}_{t+1} - \mathbf{Y}_t$ and $\mathbf{W}_{t+1}$ have the same dimensions). We also assume that $\mathbf{X}_t$ consists of $\mathbf{Y}_t - \mathbf{Y}_{t-1} - \mathbf{H}$ and a finite number of lags of this vector.

1.1 Recursive regression

Let $k \times 1$ denote the dimension of $\mathbf{Y}_t - \mathbf{Y}_{t-1}$ so that $\mathbf{W}_{t+1}$ is also a $k \times 1$ vector. Suppose that $\mathbf{X}_t$ contains $\mathbf{Y}_t - \mathbf{Y}_{t-1} - \mathbf{H}$ as well as $\ell - 1$ number of lags of $\mathbf{Y}_t - \mathbf{Y}_{t-1} - \mathbf{H}$ so that it has dimension $k\ell \times 1$. Then, $\mathbf{H}$ is a $k \times 1$ vector, $\mathbf{D}$ is a $k \times k\ell$ matrix and $\mathbf{A}$ is $k\ell \times k\ell$ matrix, $\mathbf{F}$

*Material from Takuma Habu

is a $k \times k$ matrix and $\mathbf{B}$ is a $k\ell \times k$ matrix. Then,

$$
\begin{aligned}
\mathbf{X}_{t+1} &= \begin{bmatrix} \mathbf{Y}_{t+1} - \mathbf{Y}_t - \mathbf{H} \\ \mathbf{Y}_t - \mathbf{Y}_{t-1} - \mathbf{H} \\ \vdots \\ \mathbf{Y}_{t-\ell+1} - \mathbf{Y}_{t-\ell} - \mathbf{H} \end{bmatrix}_{k\ell \times 1} \\
&= \underbrace{\begin{bmatrix} & & \mathbf{D} & & \\ \mathbb{I}_k & \mathbf{0}_k & \cdots & \cdots & \mathbf{0}_k \\ \mathbf{0}_k & \mathbb{I}_k & \mathbf{0}_k & \cdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0}_k & \cdots & \mathbf{0}_k & \mathbb{I}_k & \mathbf{0}_k \end{bmatrix}}_{=\mathbf{A}} \underbrace{\begin{bmatrix} \mathbf{Y}_t - \mathbf{Y}_{t-1} - \mathbf{H} \\ \mathbf{Y}_{t-1} - \mathbf{Y}_{t-2} - \mathbf{H} \\ \vdots \\ \mathbf{Y}_{t-\ell} - \mathbf{Y}_{t-\ell-1} - \mathbf{H} \end{bmatrix}}_{=\mathbf{X}_t} + \underbrace{\begin{bmatrix} \mathbf{F} \\ \mathbf{0}_k \\ \vdots \\ \mathbf{0}_k \end{bmatrix}}_{=\mathbf{B}} \mathbf{W}_{t+1}.
\end{aligned} \tag{1.1}
$$

* eventually, can't identify F. Only FF'.

We first factor $\mathbf{FF}'$ so that

$$\mathbf{FF}' = \mathbf{J}\Delta\mathbf{J}',$$

where $\mathbf{J}$ is a lower triangular matrix with ones on the diagonal and Δ is diagonal (LDU decomposition). Then consider

$$\mathbf{J}^{-1}(\mathbf{Y}_{t+1} - \mathbf{Y}_t) = \mathbf{J}^{-1}\mathbf{H} + \mathbf{J}^{-1}\mathbf{D}\mathbf{X}_t + \underbrace{\mathbf{J}^{-1}\mathbf{F}\mathbf{W}_{t+1}}_{:=\mathbf{U}_{t+1}}. \tag{1.2}$$

With this transformation, the variance-covariance matrix is given by

$$
\begin{aligned}
\operatorname{Var}\left[\mathbf{J}^{-1}(\mathbf{Y}_{t+1} - \mathbf{Y}_t) | \mathbf{X}_t\right] &= \operatorname{Var}[\mathbf{U}_{t+1}] = \operatorname{Var}\left[\mathbf{J}^{-1}\mathbf{F}\mathbf{W}_{t+1}\right] \\
&= \left(\mathbf{J}^{-1}\mathbf{F}\right)\mathbb{I}\left(\mathbf{J}^{-1}\mathbf{F}\right)' = \mathbf{J}^{-1}\mathbf{FF}'\left(\mathbf{J}^{-1}\right)' \\
&= \mathbf{\Delta},
\end{aligned}
$$

where we recall that $\mathbf{\Delta}$ is a diagonal matrix. Thus, the ith entry of $\mathbf{U}_{t+1}$ is uncorrelated with/independent of the jth entry of $\mathbf{U}_{t+1}$, $j \neq i$.

Notice also that $\mathbf{J}^{-1}$ is a lower triangular matrix with ones on the diagonal so that the left-hand side of (1.2) is given by

$$
\begin{aligned}
\mathbf{J}^{-1}\left(\mathbf{Y}_{t+1}-\mathbf{Y}_t\right) &= \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ j_{21} & 1 & 0 & \cdots & 0 \\ j_{31} & j_{32} & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ j_{\cancel{n}k1} & j_{\cancel{n}k2} & \cdots & j_{\cancel{n}k(\cancel{n}k-1)} & 1 \end{bmatrix} \begin{bmatrix} Y_{t+1}^{[1]}-Y_t^{[1]} \\ Y_{t+1}^{[2]}-Y_t^{[2]} \\ \vdots \\ Y_{t+1}^{[\cancel{n}k]}-Y_t^{[\cancel{n}k]} \end{bmatrix} \\
&= \begin{bmatrix} Y_{t+1}^{[1]}-Y_t^{[1]} \\ j_{21}\left(Y_{t+1}^{[1]}-Y_t^{[1]}\right)+Y_{t+1}^{[2]}-Y_t^{[2]} \\ j_{31}\left(Y_{t+1}^{[1]}-Y_t^{[1]}\right)+j_{32}\left(Y_{t+1}^{[2]}-Y_t^{[2]}\right)+Y_{t+1}^{[3]}-Y_t^{[3]} \\ \vdots \end{bmatrix}.
\end{aligned}
$$

The right-hands side is

$$
\begin{aligned}
\mathbf{J}^{-1}\mathbf{H}+\mathbf{J}^{-1}\mathbf{D}\mathbf{X}_t+\mathbf{U}_{t+1} &= \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ j_{21} & 1 & 0 & \cdots & 0 \\ j_{31} & j_{32} & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ j_{\cancel{n}k1} & j_{\cancel{n}k2} & \cdots & j_{\cancel{n}k(\cancel{n}k-1)} & 1 \end{bmatrix} \mathbf{H}_{k\times 1} \\
&+ \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ j_{21} & 1 & 0 & \cdots & 0 \\ j_{31} & j_{32} & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ j_{\cancel{n}k1} & j_{\cancel{n}k2} & \cdots & j_{\cancel{n}k(\cancel{n}k-1)} & 1 \end{bmatrix} \mathbf{D}_{k\times kl}\,\mathbf{X}_{kl\times 1}+\mathbf{U}_{t+1}
\end{aligned}
$$

$$
D_{k\times kl} = \begin{bmatrix} d_1 \\ \vdots \\ d_k \end{bmatrix}
$$

$$
\begin{aligned}
\Leftrightarrow\ \left[\mathbf{J}^{-1}\mathbf{H}+\mathbf{J}^{-1}\mathbf{D}\mathbf{X}_t+\mathbf{U}_{t+1}\right]^{[1]} &= h_1+\left[\mathbf{d}_1\mathbf{X}_t\right]^{[1]}+U_{t+1}^{[1]} \\
\left[\mathbf{J}^{-1}\mathbf{H}+\mathbf{J}^{-1}\mathbf{D}\mathbf{X}_t+\mathbf{U}_{t+1}\right]^{[2]} &= j_{21}h_1+h_2+j_{21}\left[\mathbf{d}_1\mathbf{X}_t\right]^{[1]}+\left[\mathbf{d}_2\mathbf{X}_t\right]^{[2]}+U_{t+1}^{[2]} \\
\left[\mathbf{J}^{-1}\mathbf{H}+\mathbf{J}^{-1}\mathbf{D}\mathbf{X}_t+\mathbf{U}_{t+1}\right]^{[3]} &= j_{31}h_1+j_{32}h_2+h_3+j_{31}\left[\mathbf{d}_1\mathbf{X}_t\right]^{[1]}+j_{32}\left[\mathbf{d}_2\mathbf{X}_t\right]^{[2]}+\left[\mathbf{d}_3\mathbf{X}_t\right]^{[3]}+U_{t+1}^{[3]} \\
\vdots &= \vdots
\end{aligned}
$$

We can rearrange (1.2) then as

$$
\begin{aligned}
Y_{t+1}^{[1]}-Y_t^{[1]} &= h_1+\left[\mathbf{d}_1\mathbf{X}_t\right]^{[1]}+U_{t+1}^{[1]}, \\
Y_{t+1}^{[2]}-Y_t^{[2]} &= -j_{21}\left(Y_{t+1}^{[1]}-Y_t^{[1]}\right)+j_{21}h_1+h_2+j_{21}\left[\mathbf{d}_1\mathbf{X}_t\right]^{[1]}+\left[\mathbf{d}_2\mathbf{X}_t\right]^{[2]}+U_{t+1}^{[2]},
\end{aligned}
$$

$$
\begin{aligned}
Y_{t+1}^{[3]} - Y_t^{[3]} &= -j_{31}\left(Y_{t+1}^{[1]} - Y_t^{[1]}\right) - j_{32}\left(Y_{t+1}^{[2]} - Y_t^{[2]}\right) \\
&\qquad + j_{31}h_1 + j_{32}h_2 + h_3 + j_{31}\left[\mathbf{d}_1\mathbf{X}_t\right]^{[1]} + j_{32}\left[\mathbf{d}_2\mathbf{X}_t\right]^{[2]} + \left[\mathbf{d}_3\mathbf{X}_t\right]^{[3]} + U_{t+1}^{[3]} \\
\vdots &= \quad \ddots
\end{aligned}
$$

which shows that the regression is recursive: the first equation determines the first entry of $Y_{t+1}^{[1]} - Y_t^{[1]}$, the second equation determines the first entry of $Y_{t+1}^{[2]} - Y_t^{[2]}$ which uses $Y_{t+1}^{[1]} - Y_t^{[1]}$, and so on. Recall that the ith entry of $\mathbf{U}_{t+1}$ is uncorrelated with/independent of the jth entries of $Y_{t+1} - Y_t$ for $j = 1, 2, \ldots, i-1$. Consequently, we can estimate each regression individually—there is no benefit in regressing jointly. That is, we run the following regressions:

$$
\begin{aligned}
Y_{t+1}^{[1]} - Y_t^{[1]} &= \left[\beta_1^{[1]}, \boldsymbol{\beta}_x^{[1]}\right] \begin{bmatrix} 1 \\ \mathbf{X}_t \end{bmatrix} + U_{t+1}^{[1]}, \\
Y_{t+1}^{[2]} - Y_t^{[2]} &= \left[\beta_1^{[2]}, \beta_2^{[2]}, \boldsymbol{\beta}_x^{[2]}\right] \begin{bmatrix} 1 \\ Y_{t+1}^{[1]} - Y_t^{[1]} \\ \mathbf{X}_t \end{bmatrix} + U_{t+1}^{[2]}, \\
Y_{t+1}^{[3]} - Y_t^{[3]} &= \left[\beta_1^{[3]}, \beta_2^{[3]}, \beta_3^{[3]}, \boldsymbol{\beta}_x^{[3]}\right] \begin{bmatrix} 1 \\ Y_{t+1}^{[1]} - Y_t^{[1]} \\ Y_{t+1}^{[2]} - Y_t^{[2]} \\ \mathbf{X}_t \end{bmatrix} + U_{t+1}^{[3]}, \\
\vdots &= \quad \ddots
\end{aligned}
$$

Then, we can back out the elements of $\mathbf{H}$ and $\mathbf{D}$ as

$$
\begin{aligned}
\beta_1^{[1]} &= h_1, \\
\boldsymbol{\beta}_x^{[1]} &= \mathbf{d}_1, \\
\beta_1^{[2]} &= j_{21}h_1 + h_2, \\
\beta_2^{[2]} &= -j_{21}, \\
\boldsymbol{\beta}_x^{[2]} &= j_{21}\mathbf{d}_1 + \mathbf{d}_2, \\
\beta_1^{[3]} &= j_{31}h_1 + j_{32}h_2 + h_3, \\
\beta_2^{[3]} &= -j_{31}, \\
\beta_3^{[3]} &= -j_{32},
\end{aligned}
$$

$$
\begin{bmatrix} 1 & & \\ j_{21} & 1 & \\ j_{31} & j_{32} & 1 \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} = \begin{bmatrix} \beta_1^{[1]} \\ \beta_1^{[2]} \\ \beta_1^{[3]} \end{bmatrix}
$$

$$
\underbrace{\begin{bmatrix} 1 & & \\ j_{21} & 1 & \\ j_{31} & j_{32} & 1 \end{bmatrix}}_{k\times k} \underbrace{D}_{k\times kl} = \underbrace{\begin{bmatrix} \beta_x^{[1]} \\ \beta_x^{[2]} \\ \beta_x^{[3]} \end{bmatrix}}_{k\times kl}
$$

$$\beta_x^{[3]} = j_{31}\mathbf{d}_1 + j_{32}\mathbf{d}_2 + \mathbf{d}_3.$$

This gives us a way to build the elements of $\mathbf{H}$, $\mathbf{D}$ as well as $\mathbf{J}^{-1}$. Since we know that $\mathbf{\Delta}$ is a diagonal matrix, the residuals for $(\hat{U}_{t+1}^{[1]})^2$, $(\hat{U}_{t+1}^{[1]})^2$, ..., gives the diagonal elements. We can then derive $\mathbf{FF}'$ as $\mathbf{J}\Delta\mathbf{J}'$. However, we cannot uniquely determine $\mathbf{F}$. One solution is $\mathbf{F} = \mathbf{J}\Delta^{1/2}$.[1] Knowing these parameters allows us to also back out $\mathbf{A}$ and $\mathbf{B}$ via (1.1).

1.2 Bayesian parameter learning

We write the ith regression formed in the way described as a scalar regression model:

$$Y_{t+1}^{[i]} - Y_t^{[i]} = \mathbf{R}_{t+1}^{[i]\prime}\boldsymbol{\beta}^{[i]} + U_{t+1}^{[i]}, \quad U_{t+1}^{[i]} \sim N\left(0, \sigma^2\right),$$

where $\mathbf{R}_{t+1}^{[i]}$ is the appropriate vector of regressors in the ith equation, and U_{t+1} is independent of $\mathbf{R}_{t+1}$. To simplify notation, we will omit the superscripts and just consider a single equation at a time. Suppose that $Y_{t+1} - Y_t$ is observed along with $\mathbf{R}_{t+1}$ as of date $t+1$ but not $\boldsymbol{\beta}$ or σ^2. Want to estimate β, σ,

We now wish to set up a learning model for parameters $\boldsymbol{\beta}$ and σ^2 in which the researcher updates his prior in every period as he receives more data. In theory, prior distribution is based on subjective assessment; here, we assume conjugate prior distributions as they are mathematically convenient (conjugate prior means that the prior and the posterior probabilities have the same distribution, with possibly different moments).

Let

$$\mathbf{Y}^t = \begin{bmatrix} Y_t - Y_{t-1} \\ Y_{t-1} - Y_{t-2} \\ \vdots \\ Y_1 - Y_0 \end{bmatrix}$$

and $\mathbf{X}_0$ denote the information that is observed as of date t.

We assume that

- the distribution of $\boldsymbol{\beta}$ conditioned on $\mathbf{Y}^t$, $\mathbf{X}_0$ and σ^2 is normal with mean $\mathbf{b}_t$ and precision matrix $\zeta\mathbf{\Lambda}_t$ where $\zeta = 1/\sigma^2$:

$$\boldsymbol{\beta}|\mathbf{Y}^t, \mathbf{X}_0, \sigma^2 \sim N\left(\mathbf{b}_t, (\zeta\mathbf{\Lambda}_t)^{-1}\right)$$

[1] $\Delta^{1/2}$ is built by taking the square root of each diagonal entry.

so that the associated density is given by

$$
\begin{aligned}
f\left(\boldsymbol{\beta}|\mathbf{Y}^{t},\mathbf{X}_0,\zeta\right) &= \frac{1}{\sqrt{(2\pi)^{r}\left|(\zeta\boldsymbol{\Lambda}_t)^{-1}\right|}}\exp\left[-\frac{1}{2}\left(\boldsymbol{\beta}-\mathbf{b}_t\right)'\left(\zeta\boldsymbol{\Lambda}_t\right)\left(\boldsymbol{\beta}-\mathbf{b}_t\right)\right] \\
&\propto \exp\left[-\frac{1}{2}\left(\boldsymbol{\beta}-\mathbf{b}_t\right)'\left(\zeta\boldsymbol{\Lambda}_t\right)\left(\boldsymbol{\beta}-\mathbf{b}_t\right)\right],
\end{aligned}
$$

where r is the dimension of $\boldsymbol{\beta}$ equal to $i+1$ where i denotes the ith regression.

- the distribution of ζ conditioned on $\mathbf{Y}^{t}$ and $\mathbf{X}_0$ follows a gamma distribution with shape and rate parameters, $\alpha = c_t/2+1$ and $\beta = d_t/2$:

$$
\zeta|\mathbf{Y}^{t},\mathbf{X}_0 \sim \Gamma\left(\underbrace{\frac{c_t}{2}+1}_{=\alpha},\ \underbrace{\frac{d_t}{2}}_{=\beta}\right),
$$

so that the associated density is given by

$$
\begin{aligned}
f\left(\zeta|\mathbf{Y}^{t},\mathbf{X}_0\right) &= \frac{\beta^{\alpha}}{\Gamma\left(\alpha\right)}\left(\zeta\right)^{\alpha-1}\exp\left[-\beta\zeta\right] \\
&= \frac{\frac{d_t}{2}^{\frac{c_t}{2}+1}}{\Gamma\left(\frac{c_t}{2}+1\right)}\zeta^{\frac{c_t}{2}}\exp\left[-\frac{d_t}{2}\zeta\right] \\
&\propto \zeta^{\frac{c_t}{2}}\exp\left[-\frac{1}{2}\zeta d_t\right].
\end{aligned}
$$

(Recall that $\Gamma\left(v/2,1/2\right)\overset{\mathrm{d}}{=}\chi^{2}\left(v\right)$. This means that $d_t\zeta$ has a chi-squared density with c_t+1 degrees of freedom.)

Updating mechanism for $\boldsymbol{\beta}$ In period t, given the observed data, the researcher updates his prior. The posterior then becomes the prior for period $t+1$. Using Bayes' rule, we can write the updating mechanism as

$$
\begin{aligned}
f\left(\underset{\downarrow\,\text{“A”}}{\boldsymbol{\beta}}|\zeta,\underset{\downarrow\,\text{“B”}}{Y_{t+1}-Y_t}\right) &= \frac{f\left(\boldsymbol{\beta},Y_{t+1}-Y_t|\zeta\right)}{f\left(Y_{t+1}-Y_t|\zeta\right)} = \frac{f\left(Y_{t+1}-Y_t|\zeta,\boldsymbol{\beta}\right)f\left(\boldsymbol{\beta}|\zeta\right)}{f\left(Y_{t+1}-Y_t|\zeta\right)} \\
&\propto f\left(Y_{t+1}-Y_t|\zeta,\boldsymbol{\beta}\right)f\left(\boldsymbol{\beta}|\zeta\right),
\end{aligned}
\tag{1.3}
$$

Bayes: $f(A|B)=\frac{f(A\cap B)}{f(B)}=\frac{f(B|A)f(A)}{f(B)}$

where we abbreviated the conditioning on $\mathbf{Y}^t$ and $\mathbf{X}_0$. Given that we are normal conjugate priors, we know that the posterior will also be normally distributed (see problem set). To obtain the mean the variance of the posterior, it suffices for us to consider the terms inside exp in the density function for normal.

The prior density for $\boldsymbol{\beta}$ in period t is proportional to:[2]

$$\exp\left[-\frac{1}{2}\left(\boldsymbol{\beta}-\mathbf{b}_t\right)'\left(\zeta\boldsymbol{\Lambda}_t\right)\left(\boldsymbol{\beta}-\mathbf{b}_t\right)\right] \propto f\left(\boldsymbol{\beta}\right).$$

To update his prior, the researcher uses Bayes' rule as in (1.3) by multiplying the prior density by the conditional density which is proportional to

$$\exp\left[-\frac{1}{2}\left(Y_{t+1}-Y_t-\mathbf{R}'_{t+1}\boldsymbol{\beta}\right)^2\zeta\right] \propto f\left(Y_{t+1}-Y_t|\boldsymbol{\beta}\right).$$

Thus, the posterior density is proportional to

$$\exp\left[-\frac{1}{2}\left(\boldsymbol{\beta}-\mathbf{b}_t\right)'\left(\zeta\boldsymbol{\Lambda}_t\right)\left(\boldsymbol{\beta}-\mathbf{b}_t\right)\right]\exp\left[-\frac{1}{2}\left(Y_{t+1}-Y_t-\mathbf{R}'_{t+1}\boldsymbol{\beta}\right)^2\zeta\right] \propto f\left(\boldsymbol{\beta}|Y_{t+1}-Y_t\right) \quad (1.4)$$

Sine the distribution is normal, we know that the posterior will also be normal (see problem set). We want to equate above with the density for the posterior, which is proportion to

$$\begin{aligned} f\left(\boldsymbol{\beta}|Y_{t+1}-Y_t\right) &\propto \exp\left[-\frac{1}{2}\left(\boldsymbol{\beta}-\mathbf{b}_{t+1}\right)'\left(\zeta\boldsymbol{\Lambda}_{t+1}\right)\left(\boldsymbol{\beta}-\mathbf{b}_{t+1}\right)\right] \\ &= \exp\left[-\frac{1}{2}\zeta\left(\boldsymbol{\beta}'\boldsymbol{\Lambda}_{t+1}\boldsymbol{\beta}-2\boldsymbol{\beta}'\boldsymbol{\Lambda}_{t+1}\mathbf{b}_{t+1}+\mathbf{b}'_{t+1}\boldsymbol{\Lambda}_{t+1}\mathbf{b}_{t+1}\right)\right] \end{aligned}$$

Let us simplify the expression in (1.4):

$$\begin{aligned} &\exp\left[-\frac{1}{2}\left(\boldsymbol{\beta}-\mathbf{b}_t\right)'\left(\zeta\boldsymbol{\Lambda}_t\right)\left(\boldsymbol{\beta}-\mathbf{b}_t\right)\right]\exp\left[-\frac{1}{2}\left(Y_{t+1}-Y_t-\mathbf{R}'_{t+1}\boldsymbol{\beta}\right)^2\zeta\right] \\ &= \exp\left[-\frac{1}{2}\zeta\left(\boldsymbol{\beta}'\boldsymbol{\Lambda}_t\boldsymbol{\beta}-2\boldsymbol{\beta}'\boldsymbol{\Lambda}_t\mathbf{b}_t+\mathbf{b}'_t\boldsymbol{\Lambda}_t\mathbf{b}_t\right)\right] \\ &\quad\times\exp\left[-\frac{1}{2}\zeta\left(\left(Y_{t+1}-Y_t\right)^2-2\left(Y_{t+1}-Y_t\right)\left(\mathbf{R}'_{t+1}\boldsymbol{\beta}\right)+\left(\mathbf{R}'_{t+1}\boldsymbol{\beta}\right)^2\right)\right] \\ &= \exp\left[-\frac{1}{2}\zeta\left(\left[\boldsymbol{\beta}'\boldsymbol{\Lambda}_t\boldsymbol{\beta}+\left(\mathbf{R}'_{t+1}\boldsymbol{\beta}\right)^2\right]\right)\right] \end{aligned}$$

[2] Conditioning on σ^2 is important as it allows us to use this method which is much simpler than what we did in problem set 2.

$$
\begin{aligned}
&\times \exp\left[-\frac{1}{2}\zeta\left(-2\left[\boldsymbol{\beta}'\boldsymbol{\Lambda}_t\mathbf{b}_t + (Y_{t+1} - Y_t)\left(\mathbf{R}'_{t+1}\boldsymbol{\beta}\right)\right]\right)\right] \\
&\times \exp\left[-\frac{1}{2}\zeta\left[\mathbf{b}'_t\boldsymbol{\Lambda}_t\mathbf{b}_t + (Y_{t+1} - Y_t)^2\right]\right].
\end{aligned}
$$

We first equate the quadratic terms in $\boldsymbol{\beta}$:

$$
\begin{aligned}
\boldsymbol{\beta}'\boldsymbol{\Lambda}_{t+1}\boldsymbol{\beta} &= \boldsymbol{\beta}'\boldsymbol{\Lambda}_t\boldsymbol{\beta} + \left(\mathbf{R}'_{t+1}\boldsymbol{\beta}\right)^2 \\
&= \boldsymbol{\beta}'\boldsymbol{\Lambda}_t\boldsymbol{\beta} + \left(\boldsymbol{\beta}'\mathbf{R}_{t+1}\right)\left(\boldsymbol{\beta}'\mathbf{R}_{t+1}\right)' \\
&= \boldsymbol{\beta}'\boldsymbol{\Lambda}_t\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{R}_{t+1}\mathbf{R}'_{t+1}\boldsymbol{\beta} \\
&= \boldsymbol{\beta}'\underbrace{\left(\boldsymbol{\Lambda}_t + \mathbf{R}_{t+1}\mathbf{R}'_{t+1}\right)}_{=\boldsymbol{\Lambda}_{t+1}}\boldsymbol{\beta}.
\end{aligned}
$$

We then equate the coefficients on $\boldsymbol{\beta}$:

$$
\begin{aligned}
-2\boldsymbol{\beta}'\boldsymbol{\Lambda}_{t+1}\mathbf{b}_{t+1} &= -2\left[\boldsymbol{\beta}'\boldsymbol{\Lambda}_t\mathbf{b}_t + (Y_{t+1} - Y_t)\left(\mathbf{R}'_{t+1}\boldsymbol{\beta}\right)\right] \\
&= -2\boldsymbol{\beta}'\underbrace{\left[\boldsymbol{\Lambda}_t\mathbf{b}_t + \mathbf{R}_{t+1}\left(Y_{t+1} - Y_t\right)\right]}_{=\boldsymbol{\Lambda}_{t+1}\mathbf{b}_{t+1}}.
\end{aligned}
$$

This gives us the updating equations:[3]

$$
\begin{aligned}
\boldsymbol{\Lambda}_{t+1} &= \boldsymbol{\Lambda}_t + \mathbf{R}_{t+1}\mathbf{R}'_{t+1}, \\
\boldsymbol{\Lambda}_{t+1}\mathbf{b}_{t+1} &= \boldsymbol{\Lambda}_t\mathbf{b}_t + \mathbf{R}_{t+1}\left(Y_{t+1} - Y_t\right)
\end{aligned}
$$

so that the precision of the posterior is given by $\zeta\boldsymbol{\Lambda}_{t+1}$, and the mean of the posterior is given by solving the second equation.

Since $\boldsymbol{\Lambda}_{t+1} - \boldsymbol{\Lambda}_t = \mathbf{R}_{t+1}\mathbf{R}'_{t+1} \geq 0$ (i.e. the difference is positive semi-definite), we see that precision is increasing in the additional data the researcher receives. Together,

$$
\boldsymbol{\beta}_{t+1}|\mathbf{Y}^t, \mathbf{X}_0, \sigma^2 \sim N\left(\mathbf{b}_{t+1}, (\zeta\boldsymbol{\Lambda}_{t+1})^{-1}\right).
$$

Updating mechanism for ζ So far, we have conditioned on σ^2 (i.e. ζ^{-1}). We also wish to build an updating mechanism for σ^2.

[3] Take as given that we do not need that $\mathbf{b}'_{t+1}\boldsymbol{\Lambda}_{t+1}\mathbf{b}_{t+1} = \mathbf{b}'_t\boldsymbol{\Lambda}_t\mathbf{b}_t + (Y_{t+1} - Y_t)^2$; i.e. it suffices to verify that coefficients on $\boldsymbol{\beta}$'s coincide.

Using Bayes' rule:

$$
\begin{aligned}
f\left(\zeta|\mathbf{Y}^{t+1},\mathbf{X}_0\right) &\propto f\left(Y^{t+1}-Y^t|\zeta,\mathbf{Y}^t,\mathbf{X}_0\right) f\left(\zeta|\mathbf{Y}^t,\mathbf{X}_0\right)\\
&\propto f\left(\zeta|\mathbf{Y}^t,\mathbf{X}_0\right)\int f\left(Y^{t+1}-Y^t|\zeta,\boldsymbol{\beta},\mathbf{Y}^t,\mathbf{X}_0\right) f\left(\boldsymbol{\beta}|\zeta,\mathbf{Y}^t,\mathbf{X}_0\right) d\boldsymbol{\beta}
\end{aligned}
$$

We have:

$$
\begin{aligned}
&\int f\left(Y^{t+1}-Y^t|\zeta,\boldsymbol{\beta},\mathbf{Y}^t,\mathbf{X}_0\right) f\left(\boldsymbol{\beta}|\zeta\mathbf{Y}^{t+1},\mathbf{X}_0\right) d\boldsymbol{\beta}\\
&= \frac{\zeta\sqrt{\mathrm{Det}\left(\boldsymbol{\Lambda}_t\right)}}{\sqrt{2\pi}^{r+1}}\int \exp\left[-\frac{1}{2}\left(Y_{t+1}-Y_t-\mathbf{R}_{t+1}'\boldsymbol{\beta}\right)^2\zeta\right]\exp\left[-\frac{1}{2}\left(\boldsymbol{\beta}-\mathbf{b}_t\right)'\left(\zeta\boldsymbol{\Lambda}_t\right)\left(\boldsymbol{\beta}-\mathbf{b}_t\right)\right] d\boldsymbol{\beta}\\
&\propto \zeta\exp\left(-\frac{1}{2}\left[\left(Y_{t+1}-Y_t\right)^2+\mathbf{b}_t'\boldsymbol{\Lambda}_t\mathbf{b}_t-\underline{\mathbf{R}_{t+1}\left(Y_{t+1}-Y_t\right)-\boldsymbol{\Lambda}_t}\mathbf{b}_t\right]\zeta\right)\\
&\propto \zeta\exp\left(-\frac{1}{2}\left[\left(Y_{t+1}-Y_t\right)^2+\mathbf{b}_t'\boldsymbol{\Lambda}_t\mathbf{b}_t-\mathbf{b}_{t+1}'\boldsymbol{\Lambda}_{t+1}\mathbf{b}_{t+1}\right]\zeta\right)
\end{aligned}
$$

complete the square.

As the "square" is:

$$
\begin{aligned}
-\frac{1}{2}\boldsymbol{\beta}'\zeta\left[\boldsymbol{\Lambda}_t+\mathbf{R}_{t+1}\mathbf{R}_{t+1}'\right]\boldsymbol{\beta}+2\boldsymbol{\beta}\zeta\left[\boldsymbol{\Lambda}_t\mathbf{b}_t+\mathbf{R}_{t+1}\left(Y_{t+1}-Y_t\right)\right]+\ldots\\
=-\frac{1}{2}\boldsymbol{\beta}'\zeta\boldsymbol{\Lambda}_{t+1}\boldsymbol{\beta}+2\boldsymbol{\beta}\zeta\boldsymbol{\Lambda}_{t+1}\mathbf{b}_{t+1}+\ldots
\end{aligned}
$$

Hence:

$$
\begin{aligned}
f\left(\zeta|\mathbf{Y}^{t+1},\mathbf{X}_0\right) &\propto \zeta^{\frac{c_t}{2}}\exp\left[-\frac{1}{2}\zeta d_t\right]\times\zeta\exp\left(-\frac{1}{2}\left[\left(Y_{t+1}-Y_t\right)^2+\mathbf{b}_t'\boldsymbol{\Lambda}_t\mathbf{b}_t-\mathbf{b}_{t+1}'\boldsymbol{\Lambda}_{t+1}\mathbf{b}_{t+1}\right]\zeta\right)\\
&\propto \zeta^{\frac{c_t+1}{2}}\exp\left[-\frac{1}{2}\zeta\left(\left(Y_{t+1}-Y_t\right)^2+\mathbf{b}_t'\boldsymbol{\Lambda}_t\mathbf{b}_t-\mathbf{b}_{t+1}'\boldsymbol{\Lambda}_{t+1}\mathbf{b}_{t+1}+d_t\right)\right]
\end{aligned}
$$

The implied density for ζ conditioned on time $t+1$ information has the following form:

$$
\begin{aligned}
c_{t+1} &= c_t+1,\\
d_{t+1} &= \left(Y_{t+1}-Y_t\right)^2-\mathbf{b}_{t+1}'\boldsymbol{\Lambda}_{t+1}\mathbf{b}_{t+1}+\mathbf{b}_t'\boldsymbol{\Lambda}_t\mathbf{b}_t+d_t.
\end{aligned}
$$

The density for ζ conditioned on $\mathbf{Y}^{t+1}$ and $\mathbf{X}_0$ has a gamma density:

$$
\propto \zeta^{\frac{c_t+1}{2}}\exp\left[-\frac{1}{2}\zeta d_{t+1}\right].
$$

Equivalently,

$$\zeta_{t+1}|\mathbf{Y}^{t+1}, \mathbf{X}_0 \sim \text{Gamma}\left(\frac{c_{t+1}}{2} + 1, \frac{d_{t+1}}{2}\right).$$

Remark 1. If we set $\mathbf{\Lambda}_0 = \mathbf{0}$, then the specification of $\mathbf{b}_0$ is inconsequential and $\mathbf{b}_{t+1}$ becomes the standard least squares estimator. Similarly, for ζ, we can set $c_0 = -2$ and $d_0 = 0$, which gives that d_{t+1} is the sum of squared regression residuals. This is an example of setting uninformative priors, often called "improper priors" since they do not integrate to unity. In the case of $\mathbf{b}_{t+1}$, the implied normal distribution will be proper after we accumulate enough observations to make $\mathbf{\Lambda}_{t+1}$ become nonsingular at which point $\mathbf{b}_{t+1}$ becomes uniquely identified.

Equivalently,

$$\zeta_{t+1}|\mathbf{Y}^{t+1}, \mathbf{X}_0 \sim \text{Gamma}\left(\frac{c_{t+1}}{2} + 1, \frac{d_{t+1}}{2}\right).$$

Remark 1. If we set $\mathbf{\Lambda}_0 = \mathbf{0}$, then the specification of $\mathbf{b}_0$ is inconsequential and $\mathbf{b}_{t+1}$ becomes the standard least squares estimator. Similarly, for ζ, we can set $c_0 = -2$ and $d_0 = 0$, which gives that d_{t+1} is the sum of squared regression residuals. This is an example of setting uninformative priors, often called "improper priors" since they do not integrate to unity. In the case of $\mathbf{b}_{t+1}$, the implied normal distribution will be proper after we accumulate enough observations to make $\mathbf{\Lambda}_{t+1}$ become nonsingular at which point $\mathbf{b}_{t+1}$ becomes uniquely identified.